(1) $\nabla^2 f(x) = A^T A$ 令 $M = \max eig(A^T A)$，则 M 是满足 $\nabla^2 f(x) \le MI$ 的最小值.

$$\nabla f(x) = \nabla f(y) + \nabla^2 f(y)^T (x-y)$$

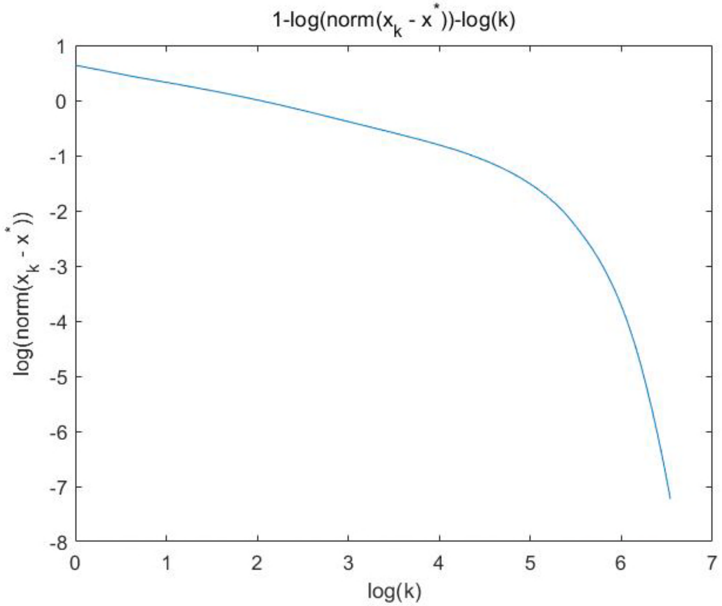
$$\Rightarrow \|\nabla f(x) - \nabla f(y)\| = \|\nabla^2 f(y)^T (x-y)\| \le M\|x-y\|$$

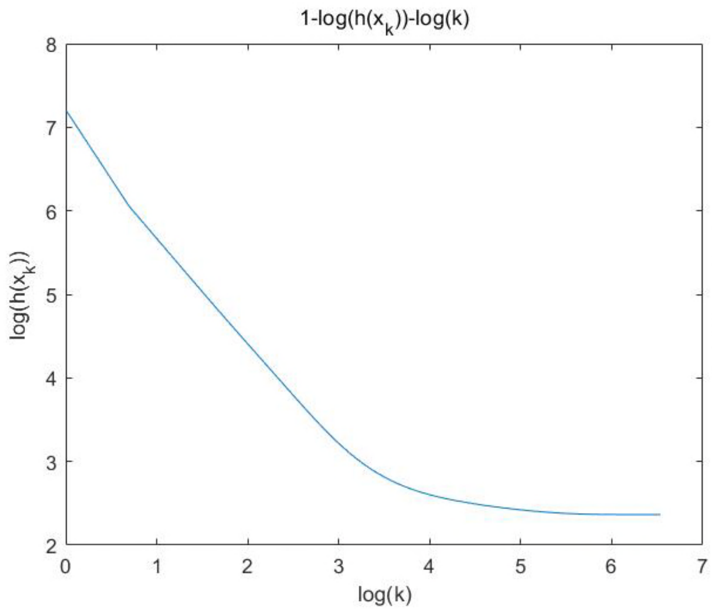
(2) $x_{k+1} = \mathrm{Prox}_{\alpha g}(x_k - \alpha \nabla f(x_k)) = S_{\alpha\gamma}(x_k - \alpha \nabla f(x_k))$，其中 $\alpha = \frac{1}{M}$, $\gamma = 1$, $\nabla f(x_k) = A^T(Ax_k - b)$

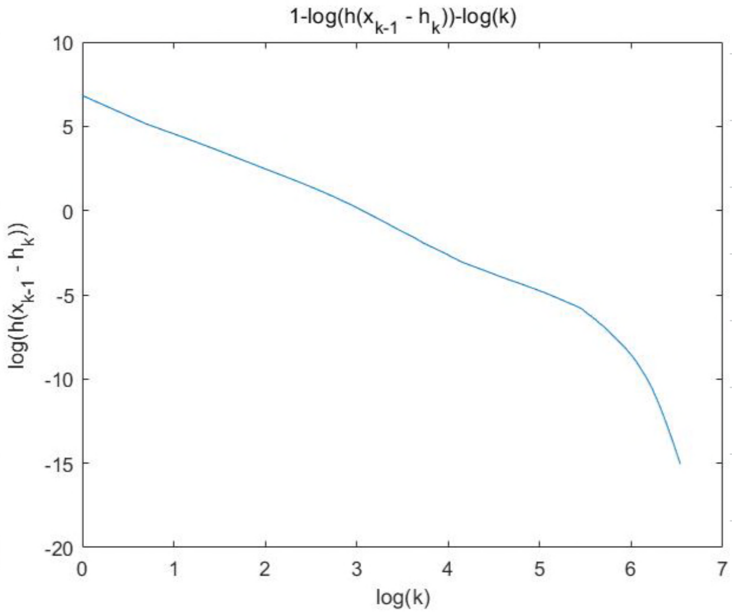
令 $[x_k]_i$ 表示 x_k 的第 i 个分量

$$[x_{k+1}]_i = \begin{cases} \left[x_k - \frac{\nabla f(x_k)}{M}\right]_i - \frac{1}{M} & \text{if } \left[x_k - \frac{\nabla f(x_k)}{M}\right]_i \ge \frac{1}{M} \\ \left[x_k - \frac{\nabla f(x_k)}{M}\right]_i + \frac{1}{M} & \text{if } \left[x_k - \frac{\nabla f(x_k)}{M}\right]_i \le -\frac{1}{M} \\ 0 & \text{otherwise} \end{cases}$$

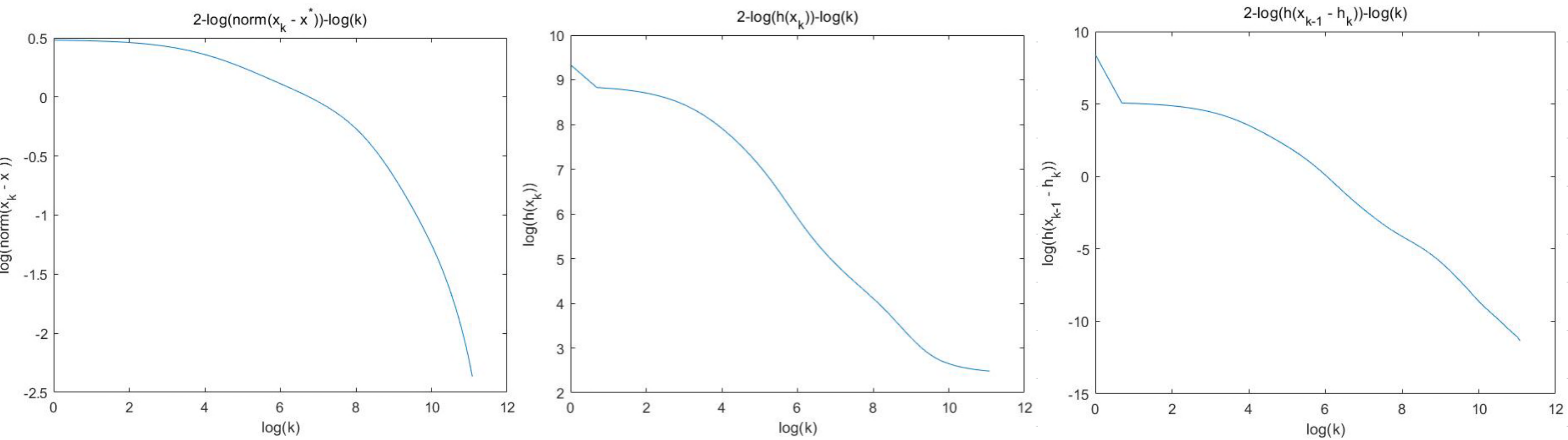
(3) A1、b1：（x^* 为 $|f(x_{k+1}) - f(x_k)| < 10^{-6}$ 时的 x_{k+1}）







A2, b2 : (x^* 为 $|f(x_{k+1}) - f(x_k)| < 10^{-6}$ 时的 x_{k+1})



(4) 令 $\Delta_k = M(x_k - x_{k+1}) = M(x_k - \mathrm{Prox}_{\alpha g}(x_k - \alpha\nabla f(x_k)))$

$\Leftrightarrow x_k - \alpha\Delta_k = \mathrm{Prox}_{\alpha g}(x_k - \alpha\nabla f(x_k)) \Leftrightarrow \Delta_k = \nabla f(x_k) + v_k,\ v_k \in \partial g(x_{k+1})$

而 $f(x_{k+1}) \le f(x_k) - \frac{1}{M}\nabla f(x_k)^T\Delta_k + \frac{1}{2M}\|\Delta_k\|^2$, $g(x_{k+1}) + \frac{1}{M}v_k^T\Delta_k \le g(x_k)$

$\Rightarrow h(x_{k+1}) \le h(x_k) - \frac{1}{2M}\|\Delta_k\|^2$

而 $h(x_{k+1}) - h(x^*) \le \nabla f(x_k)^T(x_k - x^*) + v_k^T(x_{k+1} - x^*) - \nabla f(x_k)^T(x_k - x_{k+1}) + \frac{1}{2M}\|\Delta_k\|^2$

$= \Delta_k(x_k - x^*) - \frac{1}{2M}\|\Delta_k\|^2$

$\Rightarrow h(x_k) - h(x^*) \le \frac{M\|x_0 - x^*\|^2}{2k}$

A1 对应的 M1 为 3.0401e+3，A2 对应的 M2 为 1.6384e+6，在本题题设下，更小的 M 值使得 A1 对应的情况收敛更快.